Epidemic threshold for the SIS model on random networks

Roni Parshani,[1] Shai Carmi,[1] and Shlomo Havlin[1]

[1]*Minerva Center & Department of Physics, Bar-Ilan University, Ramat Gan, Israel*

(Dated: July 16, 2010)

We derive an analytical expression for the critical infection rate r_c of the susceptible-infectious-susceptible (SIS) disease spreading model on random networks. To obtain r_c, we first calculate the probability of reinfection, π, defined as the probability of a node to reinfect the node that had earlier infected it. We then derive r_c from π using percolation theory. We show that π is governed by two effects: (*i*) The requirement from an infecting node to recover prior to its reinfection, which depends on the disease spreading parameters; and (*ii*) The competition between nodes that simultaneously try to reinfect the same ancestor, which depends on the network topology.

PACS numbers:

Diseases spread in a population as infected individuals infect other individuals with whom they are in contact. In recent years, several models have been developed to mathematically characterize the spread of diseases [1–3]. The disease spreading is best modeled as a sparse network, where the individuals are the nodes and the contacts are the links connecting them [4–6]. Two epidemic models of particular importance are the SIR (susceptible/infectious/recovered) [7] and SIS (susceptible/infectious/susceptible) [8–10]. In these models, the N network nodes are initially susceptible (they currently do not have the disease, but might become infected later), except for one randomly chosen infected node. At each time step, an infected node infects each of its neighbors with probability r. Infected nodes remain as such for τ time steps, after which they become either recovered (cannot further infect or become infected) in SIR or susceptible again in SIS. Hence, in SIS individuals can be infected multiple times.

A fundamental question in the study of epidemics is: will a disease spread throughout the population, or will it die out? The answer to this question depends on the values of the infection and recovery rates, as well as on the nature of the connections between the individuals. For the SIR model on random networks, the epidemic threshold and the critical infection rate above which the disease infects a non-zero fraction of the population were previously derived [6, 11, 12]. However, for the SIS model, the calculation of the critical infection rate is significantly more involved due to the possibility of multiple infections of the same node.

Analytical results for SIS on networks exist for a number of specific network models (e.g., a circle, a lattice, a double-layered fully connected network, etc. [13–16]). A mean-field analysis of SIS on general random networks has been proposed in [17, 18]. In this approach, master equations are written for the number of infected individuals of a given degree. To calculate the epidemic threshold, the mean-field approach assumes that an infected individual has an equal probability to infect each of its neighbors. However, as we show below, the probability of an individual to reinfect the neighbor that had previously infected it (the reinfection probability) is different from the probability to infect the other neighbors.

In this Letter, we derive an analytical expression for the SIS epidemic threshold. We find that the critical infection rate at the threshold is higher than the mean-field approximation for SIS but lower than the critical infection rate for SIR. Simulation results support the accuracy of our expression. To obtain the epidemic threshold, we first calculate the reinfection probability which depends on both the disease spreading parameters and the network topology. Then, using percolation arguments we derive the threshold from the reinfection probability.

To derive the epidemic threshold, some definitions from percolation theory are needed. In a percolation process on a network, links are removed until only a fraction p of the N network nodes remain. This is continued until a critical value, the percolation threshold p_c, is reached. For $p > p_c$, a spanning cluster of order N nodes exists, while for $p < p_c$ the network collapses into small clusters [19–23].

Disease spreading can be seen as a growing percolation process [24], in which, starting from a given seed, links are being added to the growing network with probability p. The critical infection rate in which the disease spreads throughout the network is equivalent to the percolation threshold in which a spanning cluster appears. To complete the mapping, we specify the epidemic equivalent of p, that is, the probability of a node to infect its neighbor. This probability is different from r, since an infected node i (see Fig. 1) can infect its neighbor j only as long as it is still infected, for τ time steps. Therefore, the desired probability is given by [12]:

$$p = 1 - (1 - r)^\tau. \tag{1}$$

The critical infection rate r_c for a given τ can then be obtained by substituting p_c in Eq. (1).

To find p_c, we define $\langle n_i \rangle$, the average number of susceptible nodes infected by an already infected node i. If $\langle n_i \rangle > 1$, the disease will keep on spreading until a non-zero fraction of the network is covered [25]. In SIR, only
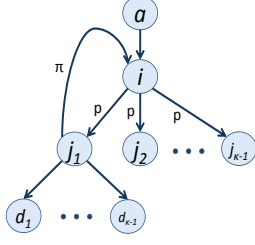
FIG. 1: A schematic illustration of the sub-network in which the probabilities of infections are calculated in the text.

the descendants of a node should be taken into account, whereas in SIS, the ancestor could also be reinfected. In terms of Fig. 1, node i, once infected, can infect not only its descendants $j_1, j_2, ...$ (in SIR) but also its ancestor a (in SIS).

The probability of a node that is reached by following a link to have degree k (one incoming link and $k - 1$ outgoing links) is $kP(k)/\langle k \rangle$ where P(k) is the degree distribution of the network nodes. Therefore, as long as the network has a tree like structure and loops are negligible [23], the average number of neighbors infected by node i for SIR is given by [25]:

$$\langle n_i \rangle_{\mathrm{SIR}} = p \sum_k \frac{kP(k)}{\langle k \rangle}(k - 1) \equiv p(\kappa - 1), \qquad (2)$$

where $\kappa - 1 \equiv [\langle k^2 \rangle / \langle k \rangle] - 1$ is the branching factor. At the epidemic threshold, $p_c(\kappa - 1) = 1$ [23]. Writing p_c in terms of r_c and τ (Eq. (1)) we obtain:

$$r_c = 1 - \left(\frac{\kappa - 2}{\kappa - 1} \right)^{1/\tau}. \qquad (3)$$

In SIS, the additional link leading to the ancestor should be incorporated into the calculation. This can be accomplished by making use of an effective branching factor, obtained by replacing $\kappa - 1$ by κ (the mean field approach). For example, consider node j_1 in Fig. 1: in addition to the $\kappa - 1$ links that are connected to the descendants $d_1, ..., d_{\kappa-1}$, it also has the link to i.

However, this approach is too simplistic: the probability of j_1 to reinfect its ancestor is different from its probability to infect its descendants. This is true since as opposed to the descendants, i is not necessarily susceptible. This is a combined outcome of two effects. First, since i must have been sick before infecting j_1 in the first place, j_1 can only reinfect i after i has recovered and became susceptible again (the "time" factor). Moreover, one of the *other* descendants of i ($j_2, ..., j_{\kappa-1}$ in Fig. 1) might reinfect i before j_1 does (the "neighbors" factor). Below, we derive an analytical expression for the probability of reinfection π, taking into account these effects and comparing it to the mean field approach.

Using π, an expression for $\langle n_i \rangle_{\mathrm{SIS}}$ immediately follows:

$$\langle n_i \rangle_{\mathrm{SIS}} = p(\kappa - 1) + \pi. \qquad (4)$$

The first term is the SIR branching factor (Eq. (2)) and the second is the SIS-specific contribution of the probability to reinfect the ancestor.

Our analysis depends on the assumption that the descendants of an infected node $d_1, ..., d_{\kappa-1}$ are susceptible and therefore the probability to infect them is p. This assumption is commonly used in epidemic models when studying disease spreading (e.g., [6, 12]), but limits the validity of the results to the region of at or below the epidemic threshold. For SIR this assumption is trivially true since reinfections are not permitted and the disease spreads directionally down a tree structure. However, for SIS where reinfections are allowed, the disease can spread down and back up the same branch, and thus the neighbors of a just infected node may already be infected. Nevertheless, since at or below the threshold the infection rate is already low, the probability that the disease has spread down and then back up is extremely small [26].

To calculate π we first obtain the probability π_t of j_1 to infect i, considering the "time" factor and ignoring the "neighbors" factor. Assume node i has been infected for s time steps before infecting j_1. Since the total lifetime of the disease is τ time steps, i remains infected for $(\tau - s)$ time steps after infecting j_1. Therefore, the total time in which j_1 is infected and i is not is $\tau - (\tau - s) = s$. The desired probability π_t is obtained by conditioning on s:

$$\pi_t = \sum_{s=1}^{\tau} \frac{(1-r)^{s-1}r}{1-(1-r)^\tau}[1-(1-r)^s] = \frac{1-(1-r)^{\tau+1}}{2-r} \quad (5)$$

where $(1-r)^{s-1}r/[1-(1-r)^\tau]$ is the probability that i infected j_1 at step s given that j_1 was eventually infected, and $[1 - (1-r)^s]$ is the probability that j_1 infected i in at most s steps.

To complete the derivation of π, we incorporate the effect of competition (which node is the first to reinfect i) between j_1 and the other descendants of i. Since i is arrived at by following a link from a, it will have on average $\kappa - 1$ descendants, $j_1, ..., j_{\kappa-1}$. In addition, i can also be reinfected by its ancestor a. Therefore [27]:

$$\pi = \pi_t \sum_{k'=0}^{\kappa-1} \binom{\kappa-1}{k'} \frac{(p\pi_t)^{k'}(1-p\pi_t)^{\kappa-1-k'}}{k'+1}. \qquad (6)$$

To see this, recall that π_t is the probability of j_1 to infect i, independently of its siblings $a, j_2, ..., j_{\kappa-1}$. Thus, $p\pi_t$ is the probability of a sibling of j_1 to be infected by i, and then reinfect i. From this, $\pi_t \binom{\kappa-1}{k'}(p\pi_t)^{k'}(1-p\pi_t)^{\kappa-1-k'}$ is the probability that *exactly* k' siblings of j_1, in addition to j_1 itself, have succeeded to reinfect i. However, after the first out of these

$k' + 1$ nodes have reinfected i, none of the other k' nodes can do so. Since all of these nodes are a-priori equivalent, the probability of j_1 to successfully reinfect i is obtained by dividing by $k' + 1$. Finally, the desired probability π is obtained by summing over all possible values of k'.

Next, we compare simulation results for the reinfection probability with the three approximations presented above: (i) p - the reinfection probability assuming it is equal to the probability to infect any other node (mean field approach); (ii) π_t - the reinfection probability taking into account only the "time" effect (Eq. (5)); (iii) π - the reinfection probability according to Eq. (6). Recall that due to the assumption explained above, the equation for π is expected to hold only at or below the epidemic threshold. Fig.2 shows that while π_t is a better estimator of the reinfection probability compared to the mean-field approximation, our Eq. (6) is superior to both and agrees with the simulation results up to the epidemic threshold (indicated with an arrow). For increasing values of τ (Figs. 2(a) and (c)) two opposing effects occur: While p by itself grows with τ, increasing the probability of reinfection, the growth of τ also reinforces the "time" and "neighbors" effects, thus attenuating the increase in π. An increase in r (Figs. 2(b) and (d)) has a similar outcome. As r increases, p increases. In addition, large values of r imply also that $j_2, ..., j_{\kappa-1}, a$ are more probable to infect i before j_1, thus decreasing the total reinfection probability π.

Our analysis has so far been general for all random networks with a prescribed degree distribution $P(k)$. Two specific models are of special interest. The first is the Erdős - Rényi (ER) network model [21, 22, 28], in which all links exist with equal probability ϕ, leading to a Poisson degree distribution $P(k) = e^{-\langle k \rangle} \langle k \rangle^k / k!$ with average degree $\langle k \rangle = (N-1)\phi$. The ER network model has become a classic model in random graph theory and was intensively studied in the past few decades. The other model is that of scale free networks (SF) [29]: networks with a broad degree distribution, usually in the form of a power-law, $P(k) \sim k^{-\gamma}$ with $\gamma > 2$. It was found that many natural networks are scale-free [29].

Using our results, the formulation of the critical infection rate for the SIS epidemic model on both ER and SF networks is straightforward. At the threshold,

$$p(r_c)(\kappa - 1) + \pi(r_c) = 1. \tag{7}$$

For ER networks, $\kappa - 1 = \langle k^2 \rangle / \langle k \rangle - 1 = \langle k \rangle$. Thus, the equation for the critical infection rate is $p(r_c) \langle k \rangle + \pi(r_c) = 1$. For SF networks, the value of $\kappa - 1$ depends on the degree exponent γ. Theoretically, κ diverges for infinite SF networks with $2 < \gamma < 3$, leading to the trivial solution $r_c = 0$. In practice, even for $\gamma < 3$ finite networks have finite degrees and thus the epidemic threshold is non-zero [30].

Simulation results for the critical infection rate r_c are presented in Fig. 3, with the theoretical values of r_c ob-
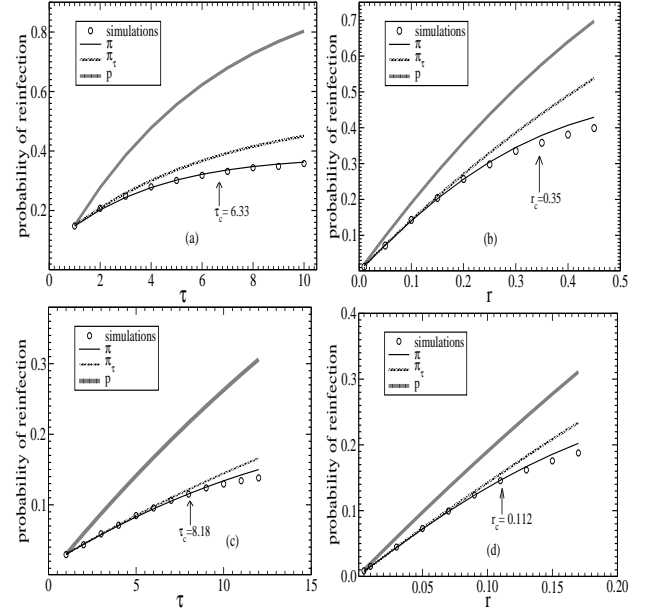

FIG. 2: Simulation and theory for the reinfection probability. Three approximations: p (dot-dashed lines), π_t (dashed lines), π (solid lines) are compared with simulation results (see text). Simulation results (circles) were obtained by recording the probability of a infected node to reinfect its ancestors, averaging over many network and disease configurations. The simulations presented are for ER networks but were validated also for SF networks. The epidemic threshold is indicated with an arrow, and is calculated from Eq. (7). (a) $\kappa = 2.1$ $r = 0.15$ and increasing values of τ. (b) $\kappa = 2.1$ $\tau = 2$ and increasing values of r (c) $\kappa = 5$ and $r = 0.03$. (d) $\kappa = 5$ and $\tau = 2$.

tained by numerically solving Eq. (7) for given τ and κ. In Fig. 3(a) we plot theory and simulations for r_c vs. τ for ER and SF networks. In Fig. 3(b) we compare our theory, the mean-field approach, and simulation results. For large values of τ, the the mean-field approach deviates by up to 10 percent, while our theory is accurate (see note [31]). The inset of Fig. 3(b) compares SIR and SIS.

The SIS epidemic model we analyzed has been limited to the case of a fixed recovery time. This is because, to calculate the epidemic threshold, we based our theory on the mapping between bond percolation and the SIR model [12, 25]. However, it has recently been shown that this mapping is valid only when the recovery time is fixed, whereas variable recovery time introduces correlations not present in the bond percolation model [32]. Nevertheless, we argue that our reinfection probability π is an upper bound for the reinfection probability with variable recovery time. This can be shown by simulations as well as by the following intuitive argument. Consider the simple case when only the time effect is taken into account and $\pi = \pi_t = [1 - (1 - r)^{(\tau+1)}]/(2 - r)$. For a fixed recovery rate, τ is a random variable, geometrically
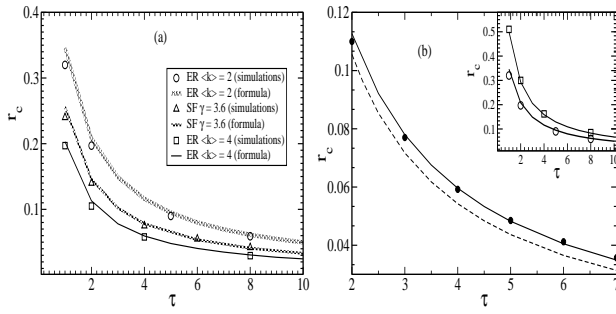
FIG. 3: Simulation results for the critical infection rate. (a) Simulation results of r_c vs. τ for ER networks with $\langle k \rangle = 2, 4$ and SF networks with $\gamma = 3.6$ (symbols), compared to theory (lines) as obtained from Eq. (7). (b) Theoretical values of r_c (Eq. (7); solid line) compared to the mean-field approach (obtained by setting $\pi = p$ in Eq. (7); dashed line) and simulation (circles) for $\langle k \rangle = 4$. The inset compares r_c (theory and simulations) between SIR (circles) and SIS (squares) with $\langle k \rangle = 2$. Theoretical values (lines) were obtained from Eq. (3) for SIR and from Eq. (7) for SIS. The epidemic threshold was calculated using a method of percolation theory: denote by S the total number of infections; r_c was identified as r for which $S \sim N$ [19, 20].

distributed with mean $\langle \tau \rangle$. Noting that $\frac{\partial^2 \pi}{\partial \tau^2} \leq 0$ for all τ, it follows from Jensen's inequality that $\langle \pi(\tau) \rangle \leq \pi(\langle \tau \rangle)$.

In summary, we analyzed the propagation of diseases on random networks according to the SIS model. The presence of reinfections in the model poses a great challenge since it can no longer be mapped into a simple branching process. We overcame this problem by calculating the reinfection probability. With this probability, we could recast the problem back into a branching process with an effective branching factor that takes into account not only the descendants of a node, but also its ancestor, weighted by the probability of reinfection. The condition for an outbreak was then found using the reinfection probability and percolation arguments.

[1] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications* (Hafner Press, New York, 1975).
[2] R. M. Anderson and R. M. May, *Infectious Diseases of Humans* (Oxford University Press, 1991).
[3] H. W. Hethcote, SIAM Review **42**, 599653 (2000).
[4] C. Moore and M. E. J. Newman, Phys. Rev. E **61**, 56785682 (2000).
[5] M. Kuperman and G. Abramson, Phys. Rev. Lett. **86**, 29092912 (2001).
[6] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 32003203 (2001).
[7] W. O. Kermack and A. G. McKendrick, Proc. R. Soc. A **115**, 700 (1927).
[8] G. H. Weiss and M. Dishon, Math Biosic **11**, 261 (1971).
[9] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases* (Griffin, London, 1975).
[10] J. D. Murray, *Mathematical Biology* (Springer-Verlag, 1993).
[11] P. Grassberger, Math. Biosci. **63**, 157 (1982).
[12] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[13] T. M. Liggett, *Stochastic Interacting systems: contact, voter and exclusion processes* (Springer, 1999).
[14] O. Diekmann, M. C. M. De Jong, and J. A. J. Metz, J. Appl. Prob. **35**, 448 (1998).
[15] F. Ball, Math. Biosci. **156**, 41 (1999).
[16] P. J. Neal, J. Appl. Prob. **45**, 513 (2008).
[17] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[18] R. P.-S. M. Boguna and A. Vespignani, Lect. Notes Comput. Sc. **127**, 625 (2003).
[19] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, 2004), 2nd ed.
[20] A. Bunde and S. Havlin, *Fractals and Disordered Systems* (Springer-Verlag, 1996), 2nd ed.
[21] P. Erdős and A. Rényi, Publ. Math. **6**, 290 (1959).
[22] B. Bollobás, *Random Graphs* (Academic Press, Orlando, 1985).
[23] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[24] Z. Alexandrowicz, Phys. Lett. A **80**, 284 (1980).
[25] N. Madar, T. Kalisky, R. Cohen, D. Ben-Avraham, and S. Havlin, Eur. Phys. J. B **38 (2)**, 269 (2004).
[26] To support the argument that at or below the threshold the descendants of an infected node are susceptible, we directly measured, by computer simulations the probability that a neighbor of an infected node is susceptible and the results show that at or below the epidemic threshold, this probability is indeed very close to 1.
[27] More accurately, one needs to distinguish between a and $j_2, ..., j_{\kappa-1}$ since a had originally infected i and is thus expected to be infected with higher probability. However, simulations show that the change in π due to this effect is negligible.
[28] P. Erdős and A. Rényi, A. Publ. Math. Inst. Hung. Acad. Sci. **5**, 1760 (1960).
[29] A.-L. Barabasi and R. Albert, Science **286**, 509 (1999).
[30] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 035108 (2002).
[31] The difference between our theory and the MF approach is nicely demonstrated using the percolation threshold, p_c. For the MF approach $p_c = 1/\kappa$, and thus for a given κ, is independent of τ, while according to our theory $p_c = [1 - \pi(\tau)]/(\kappa - 1)$ depends on τ. For the parameters used in Fig.3(b), the MF approach yields $p_c(\langle k \rangle = 4) = 0.2$ for any τ while for our theory $p_c(\langle k \rangle = 4)$ approaches 0.22575 for large τ.
[32] E. Kenah and J. M. Robins, Phys. Rev. E **036113**, 76 (2007).